

Published in final edited form as:

Am J Prev Med. 2011 February ; 40(2): 191–198. doi:10.1016/j.amepre.2010.10.022.

Estimating the Population Impact of Preventive Interventions from Randomized Trials

Thomas D. Koepsell, MD, MPH, Douglas F. Zatzick, MD, and Frederick P. Rivara, MD, MPH Harborview Injury Prevention and Research Center (Koepsell, Zatzick, Rivara), the Department of Epidemiology (Koepsell, Rivara), the Department of Health Services (Koepsell); the Department of Psychiatry (Zatzick); the Department of Pediatrics (Rivara) University of Washington, Seattle, Washington

Abstract

Growing concern about the limited generalizability of trials of preventive interventions has led to several proposals concerning the design, reporting, and interpretation of such trials. This paper presents an epidemiologic framework that highlights three key determinants of population impact of many prevention programs: the proportion of the population at risk who would be candidates for a generic intervention in routine use, the proportion of those candidates who are actually intervened on through a specific program, and the reduction in incidence produced by that program among recipients. It then describes how the design of a prevention trial relates to estimating these quantities. Implications of the framework include: (1) reach is an attribute of a program, while external validity is an attribute of a trial, and the two should not be conflated; (2) specification of a defined target population at risk is essential in the long run and merits greater emphasis in the planning and interpretation of prevention trials; (3) with due attention to sampling frame and sampling method, the process of subject recruitment for a trial can yield key information about quantities that are important for assessing its potential population impact; and (4) exclusions during subject recruitment can be conceptually separated into intervention-driven, program-driven, and trial design-driven exclusions, which have quite different implications for trial interpretation and for estimating population impact of the intervention studied.

Introduction

From a public health standpoint, a preventive intervention aims to minimize the number of people afflicted by a disease in a defined population. Randomized trials play a key role in evaluating such interventions because of their potentially high internal validity: avoidance of bias from unmeasured confounding factors and firmer support for causal inference. Two roles have been distinguished for randomized trials in this context.¹ An *efficacy* trial seeks to determine how well the intervention *can* work under conditions that maximize its chances of having a favorable impact. An *effectiveness* trial seeks to determine how well the intervention *does* work when implemented through a certain program in real-world conditions for more broadly representative recipients. Traditionally, efficacy trials come

Address correspondence to: Thomas Koepsell, MD, Department of Epidemiology, Box 357236, University of Washington, Seattle, Washington 98195-7236. koepsell@u.washington.edu.

No financial disclosures were reported by the authors of this paper.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

first, to be followed by effectiveness trials if the efficacy-trial results are promising.² This classification of trial orientations has been extended to add pragmatic *implementation* or *dissemination* trials, which also seek to assess intervention impact on a broad scale.³

Some interventions that have performed well in efficacy trials have yielded disappointing results when applied more widely.⁴ Accordingly, several authors have advocated increased emphasis on external validity (generalizability) in the design, reporting, and interpretation of trials.^{5–11} Following this advice is complicated by the fact that many factors can influence external validity. Rothwell's review⁸ listed a diverse set of 39 factors. The RE-AIM framework has been proposed as a way to conceptualize and measure dissemination potential by considering Reach, Efficacy, Adoption, Implementation, and Maintenance.^{5,6,12} An illustration of this framework combined several indicators of each of these five dimensions into a summary index that was compared between trials on prevention of diabetes complications.¹³ RE-AIM's developers and others^{9,14} have also proposed new elements related to external validity for the CONSORT guidelines on trial reporting.¹⁵

This paper describes an epidemiologic framework that is intended to clarify concepts and define key terms involved in evaluating preventive interventions. The framework may offer a foundation for such efforts as RE-AIM in addressing intervention reach and trial generalizability, which are sometimes conflated. The framework reinforces the importance of target-population specification early in trial design. It also distinguishes three kinds of exclusions in a trial—intervention-driven, program-driven, and trial design-driven—which affect trial interpretation differently.

Population Impact of a Preventive Intervention

Model components

Consider a scenario that includes:

A health condition to be prevented, such as coronary heart disease, motor vehicle collision injury, or post-traumatic stress disorder (PTSD) following serious injury.

A defined population at risk for that condition, such as residents of a geopolitical area, enrollees in an HMO, or patients hospitalized at a certain facility for serious injury and who could later develop PTSD.

A preventive intervention, such as counseling to aid smoking cessation, low-dose aspirin, or cognitive behavioral therapy to trauma victims during hospitalization.

A program through which the intervention is implemented, involving a certain quantity and mix of resources and a certain manner of deploying them.

A candidate for the intervention is defined as a member of the population at risk who would be considered a suitable intervention recipient *in routine practice*, not just in a research context. Candidacy for the intervention may depend on several factors, including:

Disease risk—The intervention may presuppose having a certain attribute associated with elevated disease risk, such as a high cholesterol level (for coronary heart disease) or being a victim of intentional injury (for PTSD).

Intervention risk—Some people may have attributes that could increase adverse effects of the intervention, such as known intolerance to low-dose aspirin.

Other personal attributes—These may concern whether the intervention could feasibly be applied to that person, such as having a telephone (for a quit line intervention) or absence of cognitive impairment (for participation in counseling to prevent PTSD).

Table 1 shows examples of which individuals in a defined population could be regarded as candidates for different preventive interventions.

Candidacy divides the defined population at risk into two mutually exclusive and collectively exhaustive subpopulations: candidates and non-candidates. Let:

I = Incidence of health condition in full population at risk

I_c = Incidence among candidates

I_{nc} = Incidence among non-candidates

π_c = Proportion of population at risk who are candidates

By a well known relationship among disease-frequency measures [16]:

$$I = I_c \cdot \pi_c + I_{nc} \cdot (1 - \pi_c) \quad (1)$$

In words, overall incidence is a weighted average of incidence among candidates and incidence among non-candidates. The weights are the population frequency of candidacy and of non-candidacy, respectively.

Overall impact of an intervention on incidence can be defined under a potential-outcomes formulation based on the Rubin Causal Model [17–21]. Two hypothetical scenarios are envisioned here: the intervention is provided to all candidates, or the intervention is not provided to anyone. Overall incidence is designated $I(1)$ under the first scenario and $I(0)$ under the second. All other factors are held constant, including time. Actually implementing such an experiment cannot be done, in part because it is impossible to turn back the clock and study both scenarios at the same time in the same population. Nonetheless, it is useful conceptually to define population impact as $\Delta = I(0) - I(1)$: that is, the overall reduction in population incidence caused by the intervention. Various evaluation designs can be viewed as different ways of estimating Δ indirectly [21]. For example, a randomized trial provides a direct estimate of $I(1)$ in the experimental group, while incidence in the control group provides an indirect estimate of $I(0)$. Notation introduced earlier can be extended to distinguish between the two scenarios. With the intervention:

$I(1)$ = Incidence in full population at risk

$I_c(1)$ = Incidence among candidates

$I_{nc}(1)$ = Incidence among non-candidates

Without the intervention:

$I(0)$ = Incidence in full population at risk

$I_c(0)$ = Incidence among candidates

$I_{nc}(0)$ = Incidence among non-candidates

As before:

$$\begin{aligned} I(0) &= I_c(0) \cdot \pi_c + I_{nc}(0) \cdot (1 - \pi_c) \\ I(1) &= I_c(1) \cdot \pi_c + I_{nc}(1) \cdot (1 - \pi_c) \end{aligned}$$

Let $\Delta_c = I_c(0) - I_c(1)$ and $\Delta_{nc} = I_{nc}(0) - I_{nc}(1)$ denote the difference in incidence between the with-intervention and without-intervention scenarios among candidates and non-candidates, respectively. By subtraction:

$$\begin{aligned} I(0) - I(1) &= [I_c(0) - I_c(1)] \cdot \pi_c + [I_{nc}(0) - I_{nc}(1)] \cdot (1 - \pi_c) \\ \Delta &= \Delta_c \cdot \pi_c + \Delta_{nc} \cdot (1 - \pi_c) \end{aligned}$$

Interventions applied to candidates can differ in their effects on non-candidates, Δ_{nc} . Some may have “spill-over” effects. For example, vaccination can induce herd immunity, and an intervention that changes social norms about the acceptability of smoking may affect smoking behavior even among people not directly targeted. However, for many interventions, such as low-dose aspirin for heart disease prevention or counseling to prevent PTSD in trauma victims, it is reasonable to assume that $\Delta_{nc} = 0$ (i.e., that intervening on candidates only has no effect on incidence among non-candidates). When $\Delta_{nc} = 0$, then:

$$\Delta = \Delta_c \cdot \pi_c \quad (2)$$

Thus, in the absence of “spill-over” effects, the population impact (Δ) of an intervention depends on two factors: (1) the reduction in incidence among intervention candidates (Δ_c)—a measure of its *effectiveness*; and (2) the prevalence of candidacy in the population at risk (π_c)—a measure of the intervention’s *breadth of applicability*.

Factoring the population impact of an intervention into two component parts is reminiscent of factoring the population attributable fraction (also termed attributable risk) for a certain exposure into two parts: prevalence of the exposure among disease cases, and attributable fraction among the exposed, $(RR-1)/RR$ [22, 23]. Here, however, the relevant prevalence is that of candidacy for an intervention, and the hypothesized effect of intervention is to reduce incidence not to the level in non-candidates but to some other level that may be higher or lower than the incidence among non-candidates.

Population Impact of Two Hypothetic Interventions

Rose^{24–26} described two prevention strategies: a *high-risk* strategy, in which intervention is aimed selectively at those most likely to develop the target condition; and a *population* strategy, aiming intervention less selectively at much or all of the population at risk. Figures 1 and 2 concern two hypothetical interventions aimed at preventing PTSD in a certain defined population at risk: namely, patients hospitalized for trauma. The figures show graphically how Δ_c and π_c combine to determine the number of cases prevented.

In both scenarios, the defined population at risk consists of 600 hospitalized trauma patients, 150 of whom would develop PTSD in the absence of any intervention. In Figure 1, 100 such patients are deemed candidates for intervention because they are at high risk for PTSD, perhaps based on trauma mechanism. Intervention reduces their PTSD risk by 40% from 0.5 to 0.3, as indicated by the dashed line. PTSD risk in non-candidates is 0.2 and is unaffected by an intervention not directed at them. The area of each shaded rectangle—(width) \times (height) = (no. of people at risk) \times (incidence proportion among them)—is thus proportional to the number of cases among candidates or among non-candidates in the absence of intervention, each case being shown as a small circle. Reducing the incidence proportion among candidates prevents 40% of the 50 cases among candidates, or 20 cases, each shown as a light circle.

In Figure 2, a different intervention is applicable to 2/3 of all trauma victims, perhaps those who speak English. Candidates for this intervention have the same intrinsic risk of PTSD (0.25) as do non-candidates. Intervention on candidates reduces their PTSD risk 20% from 0.25 to 0.2. As for intervention #1, 20 cases are prevented.

For intervention #1: $\Delta_c \cdot \pi_c = (0.5 - 0.3) \cdot 100/600 = .033$. For intervention #2, $\Delta_c \cdot \pi_c = (0.25 - 0.20) \cdot 400/600 = .033$. Thus, each intervention would reduce the cumulative incidence of PTSD in the population as a whole from 0.25 to $0.25 - 0.033 = 0.217$. In the current hypothetical population of 600 people at risk, each intervention would prevent 20 cases of PTSD, although they would probably not be the same people.

Role of program design

For present purposes, a *program* is a vehicle through which a generic intervention is actually delivered. The same generic intervention can be implemented via any of several programs. The form that a program takes is often heavily influenced by: *the nature of the intervention and program model*, which affect level of funding, staff expertise, and time required to deliver it to each recipient; *availability of resources* in relation to those requirements; *setting characteristics*, including physical, institutional, and cultural environmental factors that affect what can feasibly be done and at what cost; and possibly interactions among these factors.⁶ Programs that implement the same generic intervention can thus differ importantly as to how they identify potential recipients, how many they can serve, and what eligibility criteria they apply.

As a result, program recipients may not necessarily be representative of the larger group of all candidates for the intervention. In particular, program recipients may be skewed as to:

Level of need—Programs may seek out individuals at highest risk for the condition, or those in whom it would be most severe. For example, when vaccine against H1N1 influenza was scarce in 2009, many state and local vaccination programs prioritized pregnant women.

Accessibility—Programs may preferentially enroll people who are easiest or less costly to access, such as those who reside near program headquarters.

Likelihood of intervention impact—Programs may prioritize individuals in whom the intervention may have its greatest effect, such as those most likely to be compliant.

Ability to pay—Programs that charge recipients for service may preferentially target people who can cover those costs themselves or through insurance.

Occasionally when demand exceeds supply, a program may use a lottery or the equivalent to select who receives service. For example, when telephone-call volume to a California smoking quit line exceeded program capacity, random chance was used to decide which callers would receive immediate service and which would be advised to call back later.²⁷ These unusual examples are important because they show that sometimes it is possible for program recipients to be formally representative of a larger program of intervention candidates.

In the population context, a program's process of subject enrollment subdivides intervention candidates into program recipients and non-recipients (Figure 3). As before, overall incidence of the condition in intervention candidates (I_c) can be expressed as a weighted average of incidence in program recipients (I_r) and in non-recipients (I_{nr}), where $\pi_{r|c}$ denotes the proportion of candidates who receive the program:

$$I_c = I_r \cdot \pi_{r|c} + I_{nr} \cdot (1 - \pi_{r|c}) \quad (3)$$

Also following earlier arguments, I_c can be envisioned with the program or without it, and it can be shown that, in the absence of spill-over effects:

$$\Delta_c = \Delta_r \cdot \pi_{r|c} \quad (4)$$

Lastly, substituting (4) into (2) gives:

$$\Delta = \pi_c \cdot \pi_{r|c} \cdot \Delta_r \quad (5)$$

Overall population impact is thus the product of (1) the prevalence of candidacy for the intervention in the population at risk, (2) the proportion of candidates who receive the intervention under a particular program, and (3) the amount by which incidence is reduced among program recipients.

Role of Trial Design

Consider a two-arm, parallel-groups trial comparing a prevention program (experimental arm) with no program (control arm). Say that n_E subjects are randomized to the experimental arm and n_C to the control arm, for a total of $n = n_E + n_C$ subjects and that the subscript $i = 1 \dots n$ indexes study subjects. Let E be the set of i -values that identify subjects in the experimental group and C the set of i -values for subjects in the control group. Lastly, let Y_i^* be the true outcome status of subject i , without measurement error.

The potential-outcomes formulation can be used again as the basis for defining the internal validity of such a trial. Each trial participant has two potential outcomes: $Y^*(1)$ if assigned to the experimental arm and $Y^*(0)$ if assigned to the control arm. The average effect of intervention among subjects who are assigned to the experimental arm would be:

$$\Delta_r^* = \frac{\sum_{i \in E} [Y_i^*(0) - Y_i^*(1)]}{n_E} = \left(\frac{\sum_{i \in E} Y_i^*(0)}{n_E} \right) - \left(\frac{\sum_{i \in E} Y_i^*(1)}{n_E} \right)$$

Only $Y_i^*(1)$ is actually available for measurement, not to mention possible measurement error. Nonetheless, Δ_r^* is useful conceptually to specify what the trial is designed to estimate. The observed trial result is:

$$\Delta_r = \left(\frac{\sum_{i \in C} Y_i(0)}{n_C} \right) - \left(\frac{\sum_{i \in E} Y_i(1)}{n_E} \right)$$

The first term shows that the experience of the control group serves as the basis for estimating how the experimental group would have fared if they had instead been subject to the control condition. The trial's internal validity can be defined as the degree to which $\Delta_r^* = \Delta_r$. Comparing expressions for Δ_r^* and Δ_r reveals two main threats to validity. First,

outcomes in control-group subjects may not accurately reflect the counterfactual outcomes of experimental-group subjects had they been assigned instead to the control condition. Any such discrepancy could reflect selection bias (e.g., corruption of randomization) or chance. Second, subjects' observed outcome status may not equal their true outcome status, reflecting measurement error.

Fundamentally, internal validity concerns whether a trial correctly estimates the effect of an intervention on the intervention group actually studied. Thus, even a trial that is limited to highly selected volunteers, who may be further filtered through a complex set of exclusions, can still have high internal validity.

Assessing a trial's external validity requires specifying a defined population of trial nonparticipants and gauging the degree to which trial results (specifically, its estimate of Δ_r) would apply to them. Two such defined populations of special interest would be subgroups of the overall defined population at risk for the target condition: (1) all intervention candidates; or (2) all individuals who could receive the intervention program under study, given the trial's eligibility criteria.

Efficacy trials recruit participants in such a way that it is highly likely that $\Delta_r > \Delta_c$. To determine whether the intervention *can* work under highly favorable circumstances, participants are chosen as "good" candidates for it. Features of "goodness" may include being highly compliant, being unlikely to drop out, being unlikely to experience adverse effects of the program, and so on. If these selection criteria achieve their intended goal, it is to be expected that program effects seen in efficacy trials would generally exceed the effects seen if the same intervention approach were applied to all candidates. On the other hand, efficacy trial participants may be representative of a more restricted subpopulation of intervention candidates, defined as those who met the efficacy trial's eligibility criteria. In terms of equation (5), a higher value of Δ_r is offset by a lower value of π_{rlc} .

In contrast, effectiveness trials use less restrictive eligibility criteria, such that $\Delta_r \approx \Delta_c$ more closely. In some cases, an effectiveness trial can seek to recruit subjects literally at random from among a defined population of intervention candidates (e.g., the Multicentre Aneurysm Screening Study trial of early detection and elective repair of abdominal aortic aneurysms in British men, which recruited subjects from family doctor lists and Health Authority lists with few exclusions).²⁸

Either a narrowly applied or a broadly applied intervention can be studied with either an efficacy-trial or an effectiveness-trial orientation. For example, a broad-reach prevention strategy for heart disease might be low-dose aspirin. Candidates would include almost everyone, except a few with known intolerance to low-dose aspirin. A narrow-reach strategy might be statins for adults whose fasting cholesterol exceeds a certain threshold. Candidates would include only individuals with an elevated fasting cholesterol. Candidates for low-dose aspirin would almost certainly outnumber candidates for statins.

An efficacy trial of statins would likely recruit highly compliant subjects, perhaps chosen after a run-in phase, and subjects unlikely to drop out. An effectiveness trial of statins might draw participants from several primary-care practices, with few exclusions. (Many British trials involving multiple primary-care practices have this flavor.) An efficacy trial of low-dose aspirin is exemplified by the Physicians Health Study²⁹: participants were themselves physicians who demonstrated compliance during a run-in phase. An effectiveness trial of low-dose aspirin could draw participants from multiple primary-care sites and have much more inclusive eligibility criteria.

Frequency and Types of Exclusions during Subject Recruitment

Although the main goal of a prevention trial is typically to estimate Δ_r , if the trial is viewed in a population context, the process of recruiting trial participants may yield important information about π_c and $\pi_{r|c}$ as well. The CONSORT guidelines mandate that trial reports include a flow diagram that shows how subjects were identified from a larger pool.¹⁵ However, the sample flow diagram in the current CONSORT guidelines suggests starting with “Assessed for eligibility,” then showing how many were excluded due to “Not meeting inclusion criteria.” The model presented here suggests that it may be of value to disaggregate “Not meeting inclusion criteria” into separate classes of exclusions that have quite different implications for interpreting trial results. In particular, three types of exclusions may be of special interest:

Intervention-driven exclusions disqualify subjects because they would not be candidates for the generic intervention under any program model, inside or outside the research context. For example, individuals being screened for a trial of statins would be ineligible under an intervention-driven exclusion if their fasting cholesterol value was not elevated. The frequency of intervention-driven exclusions provides information about π_c .

Program-driven exclusions disqualify subjects because of features specific to how the generic intervention was implemented in the program under study, even though these excluded individuals could be candidates for the generic intervention in routine practice. For example, subjects who fail a run-in phase might be excluded from participation in a trial of statins but would still be candidates to receive a statin in regular medical care. The frequency of program-driven exclusions provides information about $\pi_{r|c}$. Excluding a large number of subjects for program-driven reasons signals an efficacy-trial orientation, which should raise questions about whether trial results can be safely extrapolated to all candidates for the generic intervention involved—(i.e., whether $\Delta_r \approx \Delta_c$). Excluding only a few subjects for program-driven reasons signals an effectiveness-trial orientation, under which the $\Delta_r \approx \Delta_c$ approximation may hold more closely.

Trial design-driven exclusions disqualify subjects for reasons peculiar to requirements of the research study, even if these individuals would otherwise be eligible to receive the intervention under the same program model outside the research context. For example, individuals with a low literacy level that would preclude their completion of a key outcome-measurement instrument might be excluded from an efficacy trial, even though they would be part of the program’s target population otherwise.

A separate paper illustrates how a simpler version of this categorization scheme was used to compare two PTSD prevention programs.³⁰ Admittedly, it may not always be obvious how to classify a particular exclusion: for example, subject refusal may reflect unwillingness to receive the intervention under study (an intervention-driven exclusion) or unwillingness to shoulder the respondent burden posed by a lengthy set of outcome-assessment instruments (a trial design-driven exclusion). In time, this ambiguity may be lessened by encouraging trialists to ask follow-up questions about the reasons for refusal. Sensitivity analysis can also be applied, formally or informally, to gauge the effect of alternative classifications.

Discussion

A contribution of the epidemiologic framework described here is that it may help to put such commendable efforts as RE-AIM on a firmer footing. Under the framework, the public health impact of a prevention program can be seen to depend on both the proportion of the population at risk who are candidates for a certain generic intervention, as well as on the proportion of candidates who actually receive the intervention via the program. The latter

($\pi_{r|c}$ in the notation used here) corresponds more closely to the concept of *reach* in the RE-AIM model.⁵ But comparing the reach of alternative prevention programs aimed at the same condition may not be very meaningful if the generic interventions differ. In any case, reach is an attribute of a program, not of a trial. Generalizability (external validity), in contrast, is an attribute of a trial, and the two should not be conflated. Moreover, the algebraic relationships that follow from the framework may help guide development and refinement of measures of reach and of generalizability. A recent attempt to operationalize the RE-AIM framework set forth a summary score, defined as the mean of four dimension-specific subscores, each involving a custom arithmetic combination of several items.¹³ The authors suggested that future studies experiment with different ways of combining relevant information. The present work may be a first step in that direction.

Under this framework, there is no intrinsic competition between the internal and external validity of a trial. An effectiveness trial may be just as internally valid as an efficacy trial, as long as it yields a correct estimate of the effect of the intervention under study *among trial participants*. Efficacy trials and effectiveness trials aim to estimate the effect of an intervention on different groups. Neither kind of trial should be faulted for failing to estimate accurately the effect of intervention on those who would be studied in the other kind of trial. One corollary is that effectiveness trials can be designed rigorously and should be held to high standards. Specification of a defined population at risk is a key feature of this framework. In the current PTSD-prevention examples, a registry of hospitalized trauma patients could provide a convenient way to identify a defined population.³¹ Enrollees in a certain insurance plan may also provide a good starting point. Yet many trials are conducted on convenience samples of volunteers. Such trials arguably can tell us how well a prevention program works in some collection of people, but they provide little basis for estimating the proportion of any defined population that would be candidates for the intervention or be reached by a certain program that implements it. They can also leave much uncertainty about how large a reduction in incidence to expect. Longford³² has noted that it is rarely feasible to enroll trial participants strictly at random from a defined population. Still, some recruitment strategies approximate that ideal more closely than others. From a public health viewpoint, the usefulness of results from a trial can be enhanced if the sampling frame and sampling method for trial participants is described.

Finally, the framework described here focuses attention on exclusions during trial recruitment. Intervention-driven exclusions can bear on the prevalence of candidacy for the intervention under study—a key factor in estimating its potential population impact. Program-driven exclusions can help readers gauge the program's penetration among candidates, as well as the degree to which program recipients are likely to be typical of all intervention candidates. Trial design-driven exclusions can be important to count separately because, even though they may have limited participation in the study program under research conditions, they may not limit participation under real-world conditions. These kinds of exclusions can guide judgments about the degree to which observed program effects are likely to reflect those to be expected in routine practice. The frequency and nature of exclusions can thus tell us where a trial fits in the longer arc of evidence development about a preventive intervention.

All models simplify reality, and that simplification imposes limitations. One possibility has been ignored: that an intervention may be applied, intentionally or not, to people outside the working definition of candidacy, as when a drug is used for off-label indications. Figure 3 subdivides a population at risk using only two branch points. A more complex model might “un-bundle” a set of eligibility criteria to create a more elaborate tree with multiple branch points (and multiple associated parameters to be estimated). Personal characteristics that underlie different exclusions may be associated with each other, which may complicate the

task of estimating the consequences of adding or removing a particular exclusion. Lastly, difference-in-incidence measures have been used to quantify intervention impact, even though ratio-of-incidence measures are also often used to report trial results. These two kinds of measures are easily reconciled if incidence in at least one group is known or estimable.

These limitations notwithstanding, it is hoped that the epidemiologic framework presented here will be useful to trial designers, to authors and editors, and to consumers of research results. It may also help to guide how quantitative assessments of population impact may influence ratings of “overall impact” when competing grant applications are reviewed for scientific merit by federal health agencies.

Acknowledgments

The authors thank Drs. Noel Weiss and Andrew Zhou for helpful comments on early drafts of this paper and an anonymous reviewer for several excellent suggestions.

This work was supported in part by grant number R49/CE000197 from the National Center for Injury Prevention and Control, the CDC and by grants R01/MH073613 and K24/MH086814 from the National Institute of Mental Health.

References

1. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 1986;15:451–474. [PubMed: 3534875]
2. Greenwald P, Cullen JW. The new emphasis in cancer control. *J Natl Cancer Inst* 1985;74:543–551. [PubMed: 3883037]
3. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290:1624–32. [PubMed: 14506122]
4. Hallfors D, Cho H, Sanchez V, Khatapoush S, Kim HM, Bauer D. Efficacy vs effectiveness trial results of an indicated “model” substance abuse program: implications for public health. *Am J Public Health* 2006;96:2254–9. [PubMed: 16809591]
5. Glasgow RE, Lichtenstein E, Marcus AC. Why don’t we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health* 2003;93:1261–7. [PubMed: 12893608]
6. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health* 2007;28:413–33. [PubMed: 17150029]
7. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval Health Prof* 2006;29:126–53. [PubMed: 16510882]
8. Rothwell PM. External validity of randomised controlled trials: “To whom do the results of this trial apply?”. *Lancet* 2005;365:82–93. [PubMed: 15639683]
9. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;297:1233–40. [PubMed: 17374817]
10. Blanco C, Olfson M, Okuda M, Nunes EV, Liu SM, Hasin DS. Generalizability of clinical trials for alcohol dependence to community samples. *Drug Alcohol Depend* 2008;98:123–128. [PubMed: 18579319]
11. Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med* 2008;168:133–5. [PubMed: 18227357]
12. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999;89:1322–7. [PubMed: 10474547]
13. Glasgow RE, Nelson CC, Strycker LA, King DK. Using RE-AIM metrics to evaluate diabetes self-management support interventions. *Am J Prev Med* 2006;30:67–73. [PubMed: 16414426]

14. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390. [PubMed: 19001484]
15. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:698–702.
16. Koepsell, TD.; Weiss, NS. *Epidemiologic methods: Studying the occurrence of illness*. New York: Oxford University Press; 2003.
17. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974;66:688–701.
18. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc* 2005;100:322–31.
19. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med* 2007;26:20–36. [PubMed: 17072897]
20. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health* 2000;21:121–45. [PubMed: 10884949]
21. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M, Mullen PD. Alternatives to the randomized controlled trial. *Am J Public Health* 2008;98:1359–66. [PubMed: 18556609]
22. Benichou J. A review of adjusted estimators of attributable risk. *Stat Methods Med Res* 2001;10:195–216. [PubMed: 11446148]
23. Benichou J. Biostatistics and epidemiology: measuring the risk attributable to an environmental or genetic factor. *C R Biol* 2007;330:281–98. [PubMed: 17502285]
24. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;6:1–8.
25. Rose, GA. *The strategy of preventive medicine*. New York: Oxford University Press; 1992.
26. Ahern J, Jones MR, Bakshis E, Galea S. Revisiting Rose: comparing the benefits and costs of population-wide and targeted interventions. *Milbank Q* 2008;86:581–600. [PubMed: 19120981]
27. Zhu SH, Anderson CM, Tedeschi GJ, Rosbrook B, Johnson CE, Byrd M, Gutierrez-Terrell E. Evidence of real-world effectiveness of a telephone quitline for smokers. *N Engl J Med* 2002;347:1087–93. [PubMed: 12362011]
28. Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RAP, Thompson SG, Walker NM. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *Lancet* 2002;360:1531–9. [PubMed: 12443589]
29. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989;321:129–35. [PubMed: 2664509]
30. Zatzick DF, Koepsell T, Rivara FP. Using target population specification, effect size, and reach to estimate and compare the population impact of two PTSD preventive interventions. *Psychiatry* 2009;72:346–59. [PubMed: 20070133]
31. Zatzick DF, Galea S. An epidemiologic approach to the development of early trauma focused intervention. *J Trauma Stress* 2007;20:401–412. [PubMed: 17721951]
32. Longford N. Selection bias and treatment heterogeneity in clinical trials. *Stat Med* 1999;18:1467–74. [PubMed: 10398285]

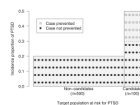


Figure 1.
Hypothetic Intervention #1 to Prevent PTSD in Trauma Victims, Illustrating the “Narrow-Reach” Strategy
PTSD, post-traumatic stress disorder

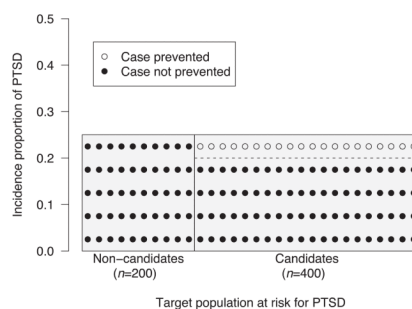


Figure 2.
Hypothetic Intervention #2 to Prevent PTSD in Trauma Victims, Illustrating the “Broad-Reach” Strategy
PTSD, post-traumatic stress disorder

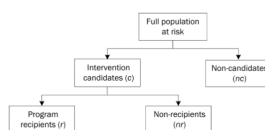


Figure 3.
Formation of Key Subpopulations from an Overall Defined Population at Risk

Table 1

Examples of candidates for different preventive interventions

| Health condition | Intervention | Candidates |
|------------------------|----------------------------------|---|
| Coronary heart disease | Smoking quit line | Smokers with a telephone |
| Coronary heart disease | Statins | Adults with high serum cholesterol and no known contraindication to statins |
| Coronary heart disease | Low-dose aspirin every other day | Adults with no known contraindications to aspirin |
| Falls in older adults | Hip pads | Ambulatory older adults |